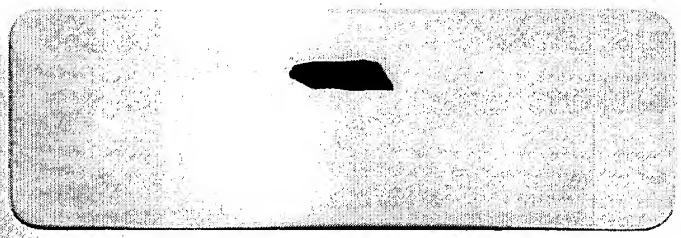
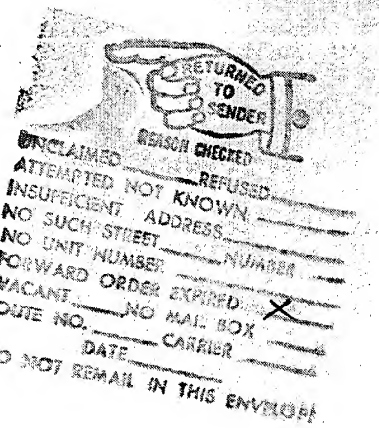
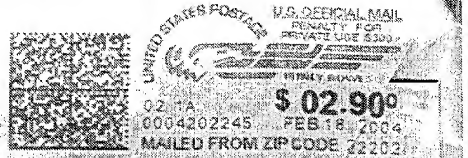
Organization 1600 Bldg./Room R3
U. S. DEPARTMENT OF COMMERCE
COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450
IF UNDELIVERABLE RETURN IN TEN DAYS

OFFICIAL BUSINESS

AN EQUAL OPPORTUNITY EMPLOYER







UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/606,977	06/28/2000	Joseph R. Byrum	16517.144/38-21(15877)B	6609

2831 7590 02/18/2004

NICOLO R. CAPOTORTO
110 BEDFORD DR.
PORT CHARLOTTE, FL 33952

EXAMINER	
ALLEN, MARIANNE P	
ART UNIT	PAPER NUMBER
1631	

DATE MAILED: 02/18/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/606,977	BYRUM, JOSEPH R.	
	Examiner	Art Unit	
	Marianne P. Allen	1631	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-7 and 20-24* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-7 and 20-24* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Applicant filed an appeal brief on 08 September 2003 and a copy of the appeal brief on 21 November 2004. Upon consideration of the record including arguments in the brief, the claims on appeal, and further review of the prior art, finality of the rejection of the last Office action (mailed 08 April 2003) is withdrawn. It is believed that this application is not ripe for appeal as all of the issues have not been developed fully on the record. New grounds of rejection are also set forth below.

Claims 1-7 and 20-24 are pending and under consideration.

Applicant is advised that the appeal brief filed refers to soybean plants in multiple places. It appears it should have referred to corn plants. Applicant is requested to edit their responses carefully to reflect the particular claims and fact pattern under consideration.

Claim Rejections - 35 USC § 101/112

Claims 2-3 and 6-7 are rejected under 35 U.S.C. 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. This is a new matter rejection.

Claims 2-3 and 6-7 have been amended to indicate that the nucleic acid molecule according to claim 1 "further comprises" an additional element. None of the portions of the specification pointed to provide support for these concepts.

Original claim 2 recited the "substantially purified nucleic acid molecule according to claim 1, wherein said nucleic acid molecule comprises a microsatellite sequence." As amended claim 2 recites the "substantially purified nucleic acid molecule according to claim 1, wherein said nucleic acid molecule further comprises a microsatellite sequence." Claim 1 is directed to a substantially purified nucleic acid molecule. In the embodiment where the nucleic acid molecule of claim 1 is the complete sequence of SEQ ID NO: 1, original claim 2 is interpreted to mean that SEQ ID NO: 1 itself contains a microsatellite sequence. This seems to be the intent of the specification (see at least page 1518-1519, bridging sentence, of the specification). Original claim 2 is not interpreted to mean that a microsatellite sequence in addition to SEQ ID NO: 1 is present in the nucleic acid molecule. This concept is embraced by amended claim 2 and the specification does not appear to contemplate this. According to this interpretation, the amended claims would fairly encompass a sequence fully containing SEQ ID NO: 1 with an additional and unrelated microsatellite sequence (claim 2), with an additional and unrelated region containing a single nucleotide polymorphism (claim 3), and with an additional and unrelated promoter or partial promoter region (claim 6). The amended claims would embrace combinations of these as well. Applicant is requested to point to basis in the specification for these specific embodiments. The examiner has only been able to find contemplation of where SEQ ID NO: 1 or a fragment thereof has these characteristics. This is not what the claims are directed to. The specification does not contemplate such nucleic acid molecules.

Claims 3 and 6-7 are considered to be new matter for the same reasons.

Applicant is requested to provide an explanation based upon the specification in support of their interpretation of the claims, either original or amended.

Claims 1-7 and 20-24 are rejected under 35 U.S.C. 101 because the claimed invention is not supported by either a specific, substantial, and credible asserted utility or a well established utility.

The sequence listing identifies SEQ ID NO: 1 as a 280 nucleotide DNA sequence with at least one wild-card nucleotide position from *Zea mays*. Table A on page 18 indicates that SEQ ID NO: 1 corresponds to clone ZM_001_A1_A01 with SEQ ID forward as ZM_001_A1_A01_T7C. These designations are not further explained.

There does not appear to be a direct assertion as to how to use SEQ ID NO: 1. There do not appear to be any particular functional characteristics of the sequence identified. While the specification generally states that SEQ ID NOS: 1-82359 encode proteins (see page 10, lines 12-16), the specification also states that SEQ ID NOS: 1-82359 are promoters (see page 11, lines 7-9) and that SEQ ID NOS: 1-82359 are markers (see page 12, lines 16-18). These are mutually exclusive classes of nucleotide sequences. For example, promoters do not encode proteins. As such, the specification does not fairly identify what SEQ ID NO: 1 is and as such, the specification cannot be considered to disclose how to use it without confirming any one of these uses or identifying an undisclosed use. Note that the specification does not disclose an open reading frame for SEQ ID NO: 1 nor is one apparent. Note that the specification does not disclose that SEQ ID NO: 1 is a repetitive sequence in *Zea mays* that has been shown to be a marker of any trait. SEQ ID NO: 1 does not appear to share significant structure with any known

marker of *Zea* mays. Note that the specification does not disclose a promoter activity for SEQ ID NO: 1 with respect to any encoded protein. SEQ ID NO: 1 does not appear to share significant structure with any known promoter. Applicant has repeatedly declined to identify which of these classes, if any, SEQ ID NO: 1 belongs to. (See at least pages 8-9, bridging paragraph of the brief.) As the functional identity of SEQ ID NO: 1 speaks to an evaluation of its utility and how to use it, applicant is being deliberately obstructive and misleading in their responses. If SEQ ID NO: 1 is or includes a promoter, then those utilities disclosed particular to promoters would be germane. However, if SEQ ID NO: 1 encodes a protein, then those utilities disclosed particular to proteins would be germane. The examiner can only conclude that applicant has not identified what for SEQ ID NO: 1 is and that the disclosure in the specification is at best misleading and at worst incorrect.

Utility of the claimed nucleic acid molecules must be evaluated as though SEQ ID NO: 1 is an uncharacterized piece of DNA.

The asserted uses for general, uncharacterized pieces of DNA have been addressed previously on the record and are summarized below.

The examiner agrees that the "The threshold of utility is not high: An invention is 'useful' under section 101 if it is capable of providing some identifiable benefit," with the proviso that the benefit be "identifiable" in the original disclosure either as a specific assertion or being readily apparent from the disclosure (i.e. well established). The examiner also agrees "the invention must have specific, i.e. not vague or unknown benefit" and "must provide a real world, i.e. practical or substantial, benefit."

It is noted that the brief states in the footnote 2 on page 6 that it "is irrelevant whether the corresponding mRNA or polypeptide have utility because Applicants are not

relying on utility of the mRNA or polypeptide to establish utility of the clamed nucleic acid molecules." The brief does not dispute that no open reading frame (ORF), no encoded protein, nor any biological activity for any encoded protein has been disclosed for SEQ ID NO: 1. Nor has SEQ ID NO: 1 been specifically identified as containing any particular promoter, polymorphism, or microsatellite marker element.

Applicant argues that the claimed nucleic acid molecules can be used to detect the presence and/or identity of polymorphisms, as hybridization probes for expression profiling, as antisense inhibitors by introduction of the claimed nucleic acid molecules into a plant or plant cell where the resulting cell or plant is to be used to screen compounds such as herbicides, to measure the level of mRNA in a sample, and as a molecular marker. The Examiner maintains that further research is required for such uses.

Use as antisense inhibitors would require further experimentation to determine the target of inhibition. These targets are not disclosed in the specification. Applicant's arguments with respect to cell based assays are not persuasive. MPEP 2107 states, "An assay that measures the presence of a material which has a stated correlation to a predisposition to the onset of a particular disease condition would also define a "real world" context of use in identifying potential candidates for preventive measures or further monitoring." The instant specification sets forth no such correlation for any condition. It is noted that this section of the MPEP goes on to state that:

> **On the other hand, the following are examples of situations that require or constitute carrying out further research to identify or reasonably confirm a "real world" context of use and, therefore, do not define "substantial utilities":**
>
> **(A) Basic research such as studying the properties of the claimed product itself or the mechanisms in which the material is involved;**
>
> **(B) A method of treating an unspecified disease or condition;**
>
> **(C) A method of assaying for or identifying a material that itself has no specific**

and/or substantial utility;
(D) A method of making a material that itself has no specific, substantial, and credible utility; and
(E) A claim to an intermediate product for use in making a final product that has no specific, substantial and credible utility.

All of these situations more closely match applicant's disclosed uses. They do not define substantial utilities.

Footnote 4 on page 10 of the brief states discusses uses of microarrays. Applicant is not claiming microarrays or collections of nucleotides and the specification does not associate any of the claimed sequences with any trait of interest. Contrary to applicant's assertions, further experimentation is required to identify a "real world use." A negative result to such a screen tells what the nucleic acid is not and cannot be used for. A positive result to such a screen requires further experimentation to determine what, if anything, such a change means. It is not an immediate benefit except in the sense to indicate that further research might yield a "real world use."

The brief on page 11 discusses gas chromatographs. MPEP 2107 in discussing research tools sets forth the following:

Some confusion can result when one attempts to label certain types of inventions as not being capable of having a specific and substantial utility based on the setting in which the invention is to be used. One example is inventions to be used in a research or laboratory setting. Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific and unquestionable utility (e.g., they are useful in analyzing compounds). An assessment that focuses on whether an invention is useful only in a research setting thus does not address whether the invention is in fact "useful" in a patent sense. Instead, Office personnel must distinguish between inventions that have a specifically identified substantial utility and inventions whose asserted utility requires further research to identify or reasonably confirm. Labels such as "research tool," "intermediate" or "for research purposes" are not helpful in determining if an applicant has identified a specific and substantial utility for the invention.

Again, further experimentation is required to use determine and confirm any of the uses set forth by applicant for the claimed nucleotide sequences.

The gas chromatograph example set forth by applicant, particularly as discussed in Footnote 5 on page 12, is not analogous to the present disclosure. A gas chromatograph is a piece of equipment designed and built for a particular use. Such equipment is fully tested, evaluated, and calibrated to ensure accurate results. Those skilled in the art use gas chromatographs to analyze both known and unknown compounds. When the compound is unknown, the results obtained are compared to results for known compounds, e.g. standards. Applicant did not design the claimed nucleotide sequences for any particular purpose. They merely isolated them. They have not tested, evaluated, or calibrated the claimed nucleotide sequences for any particular use. Sampling for the presence or absence of chlorine in a crude oil sample is not analogous to the present situation. The presence or absence of chlorine in a crude oil sample has a known meaning based upon prior research. Absent establishment of this association between presence of chlorine and destruction of catalyst, the presence or absence of chlorine in a sample would not provide any useful information to the refinery manager. Likewise, the presence or absence of any of the claimed nucleotide sequences in a sample (or polymorphisms thereof) has no meaning absent some association. Further experimentation is required to determine what that meaning or association might be.

In addition, this gas chromatograph analogy fails address applicant's own definition of the term polymorphism. The specification (page 1564, lines 1-5) defines "polymorphism" as "a variation or difference in the sequence of the gene or its flanking regions that arises in some members of a species." It follows from this definition that if there is no "variation or difference in the sequence of the gene or its flanking regions" among "members of a species," then no polymorphism exists, i.e. a polymorphism is absent, in this region of the genome. A "polymorphism" is a collective concept defined by at least two variants (or alleles) found within members of a species collectively. Thus, one detects the *presence* of a polymorphism by analyzing multiple members of the species, i.e. analyzing a population. While one can detect the absence (or presence) of a specific allele of the polymorphism in a specific individual member of the species, one cannot detect the *absence* of a polymorphism *per se* based on one individual alone. The absence of a particular allele necessarily means that a different allele is present. The specification fails to disclose a specific and substantial utility for the claimed invention in the capacity of detecting polymorphisms, because it does not disclose whether the claimed nucleic acid molecules can, in fact, be used to detect any polymorphism whatsoever. Thus, the specification leaves open the possibility that there may be no polymorphism to detect. With respect to the gas chromatograph analogy, one can only detect the absence of a compound, such as chlorine, in a sample, *if* it was already known that chlorine could, in fact, be detected by the gas chromatograph were it present in the sample.

The specification generally teaches using the claimed polynucleotides to identify a polymorphism, but fails to teach that a polymorphism could in fact be detected, or a

specific polymorphism that could be detected. The specification generally teaches using a polymorphism, detectable with the claimed nucleic acid molecules, as a molecular marker for a linked trait of interest, but fails to teach either the polymorphism or the trait of interest. The court in *Kirk* (at page 53) held:

> We do not believe that it was the intention of the statutes to require the Patent Office, the courts, or the public to play the sort of guessing game that might be involved if an applicant could satisfy the requirements of the statutes by indicating the usefulness of a claimed compound in terms of possible use so general as to be meaningless and then, after his research or that of his competitors has definitely ascertained an actual use for the compound, adducing evidence intended to show that a particular specific use would have been obvious to men skilled in the particular art to which this use relates.

The specification (page 1564, lines 1-5) defines "polymorphism" as "a variation or difference in the sequence of the gene or its flanking regions that arises in some members of a species" (emphasis added). The following pages of the specification discuss various types of sequence polymorphisms and how they are detected. It is noted that on page 1567, line 17, the specification states, "By correlating the presence or absence of it [a polymorphism] in a plant with the presence or absence of a phenotype..." Thus, the specification acknowledges that further analysis is required to determine a use for a polymorphism even assuming one is found. A change of phenotype and correlation with phenotype must be found; linkage analysis must be performed.

Even to determine whether a polymorphism exists at a specific chromosomal location requires hybridization to at least two individual chromosomes, and generally involves analyzing genomic DNA from multiple members of a species; the specification discloses no such analysis. The specification fails to disclose: 1) whether the claimed nucleic acid molecule can in fact detect a polymorphism, or even whether such a

polymorphism exists; and 2) at least one specific example of at least one of the types of polymorphisms described in the specification. The specification does not disclose any utility in this context for a nucleic acid molecule or EST that can NOT detect a polymorphism. Therefore, using the claimed invention to first determine whether or not the claimed nucleic acid molecule can, in fact, detect a polymorphism *is* to determine whether or not the claimed invention has a utility that requires detecting a polymorphism, i.e. it is "use testing" and not substantial. Since the specification fails to identify even one specific polymorphism that can be detected with the claimed nucleic acid molecule, the specification fails to show any specific correspondence between the disclosed general utility and the claimed subject matter, regardless of any specific application requiring detection of polymorphisms.

Applicant argues that the claimed nucleic acid molecules have utility as "probes for other molecules or as a source of primers." In particular, to use the claimed nucleic acid molecules to find the promoter of the corresponding gene or to initiate a chromosome walk. The argument in the brief compares the claimed invention to a microscope.

A microscope is useful for determining structure of *any* sample of interest at the macroscopic, microscopic or molecular level, depending on the type of microscope. It is a generally useful tool for a wide range of specific uses. One does not usually use a microscope to study related microscopes. In contrast, applicant argues that the claimed nucleic acid molecules are useful to detect or measure nucleic acid molecules that possess a certain level of structural relatedness to the claimed nucleic acid molecules, the level of relatedness being defined by hybridization to the claimed nucleic acid molecules.

However, the specification discloses *no* nucleic acid molecule that hybridizes with the claimed nucleic acid molecules that does *not* consist or comprise SEQ ID NO: 1 or its complement. In order for hybridization between two nucleic acid molecules to occur, they must share at least some nucleotide sequence that is fully complementary. The length of fully complementary sequence required to detect hybridization depends primarily on the stringency of the specific hybridization conditions employed, the lower the stringency the shorter the length of fully complementary sequence required. The specification also fails to disclose any hybridization conditions required for detecting nucleic acid molecules that do *not* contain the nucleotide sequence of any of SEQ ID NO: 1 or its complement (other than subsequences of SEQ ID NO: 1), in addition to failing to disclose any source for such nucleic acid molecules.

All arguments pertaining to the utility of the claimed invention with respect to studying the corresponding genomic DNA and mRNA found in maize or corn, would also apply to any homologous nucleic acid molecules found in other plant species. In so much as the specification fails to describe a specific and substantial utility for the corresponding nucleic acids in maize or corn, so does it fail to describe a specific and substantial utility for the corresponding nucleic acids in other plant species.

Applicant cites *Carl Zeiss Stiftung v. Renishaw PLC* in support of their position that utility has been established. However, this decision is with respect to a mechanical device and not a laboratory reagent or research tool. Furthermore, applicant mischaracterizes the findings in this decision. This decision concerned claim interpretation and the CAFC found that the district court had erred in their interpretation of what the claim embraced and thus what was required to establish utility. The claimed

device was found to fulfill the stated objective of mounting a stylus by the CAFC. These facts do not correspond to the instant specification.

While the specification teaches (page 1563, lines 1-7) that the claimed nucleic acid molecules "*can be used* ... to isolate molecules from other cereals" (emphasis added), the specification does not indicate that any such nucleic acid molecules *had been* obtained, nor does it describe any characteristics possessed by such nucleic acid molecules. As to whether such molecules could, in fact, be obtained, the Office can neither prove nor disprove the assertion because the Office does not have laboratory facilities. At the time the application had been filed, future experimentation on the part of one skilled in the art would have been required to determine which, if any, other plant species contained nucleic acid molecules that could have been obtained using the claimed invention, and under what experimental conditions.

With respect to using the claimed nucleic acid molecules to initiate a chromosome walk, such as to isolate a promoter of the corresponding gene, the specification fails to disclose any characteristics of the corresponding promoter, or any other promoter within "chromosome walking" distance; neither structural characteristics, by which the promoter might be identified, nor functional characteristics, by which a specific and substantial use for the promoter might be determined.

In this context, the claimed invention does not compare to a golf club, because one knows what a golf ball is and how to use the golf club to hit it, whereas the specification does not disclose or describe with particularity any known useful nucleic acid molecule that can be obtained, such as the corresponding promoter - it simply invites the skilled artisan to provide such information by further experimentation.

Even assuming *arguendo* that the corresponding promoter exists is no more guidance for its isolation, and eventual use, than knowing that a haystack contains a needle - at least one is presumed to know what the needle looks like. Also, the specification does not disclose the distance or direction one has to walk on a chromosome from the corresponding location to reach the corresponding promoter. Thus, starting the walk at the corresponding chromosomal location is no more help in identifying the promoter than is picking a specific location in a haystack to start looking for a needle when one does not know where the needle is relative to the starting location. Initiation of a chromosome walk at the corresponding chromosomal location is considered non-specific because any EST would serve the purpose for isolating an uncharacterized promoter, since any chromosomal location is expected to be linked to a promoter. The specification fails to disclose sufficient characteristics of the corresponding promoter, such as its sequence or precise location relative to the genomic location corresponding to the claimed nucleic acid molecule, to inform one of what the corresponding promoter is or when it has been isolated. For example, a nucleotide sequence is identified during the chromosome walk as a putative promoter by sequence analysis, is then subcloned into operable linkage with a reporter gene and transfected into an appropriate cell, but found not to express the reporter gene in the cells. This result could mean the putative promoter: is not truly a promoter, i.e. a false positive; is not the corresponding promoter; or is incomplete, i.e. lacked additional sequence elements required for promoter activity in the seed pod cells. Substantial utility means that "one skilled in the art can use a claimed discovery in a manner which provides some *immediate* benefit to the public," *Nelson v. Bowler*, 206 USPQ2d 881, 883 (CCPA 1980) (emphasis added). Since the specification

does not describe the corresponding promoter, or any other specific nucleic acid molecule, sufficient to inform one skilled in the art that it has been isolated, there can be no "*immediate* benefit to the public" in using the claimed nucleic acid molecule in this capacity; "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion," *Brenner* at page 696.

With respect to the "real world" value of ESTs in general (brief, pages 13-14), it is asserted that there is "no question that the public has recognized the benefits provided by the claimed subject matter, and has attributed 'real world' value to such nucleic acid molecules." It is unclear as to what evidence applicant is alluding. The evidence supplied by applicant shows that a multimillion dollar industry has arisen surrounding buying and selling EST databases and clones, not that anyone in this industry has bought or sold the claimed subject matter. It is further noted that this industry is presently in decline as evidenced by companies laying off large portions of their work force as well as moving away from EST data as their core business. It is noted that simply because a product, such as an EST sequence database or clone library, is bought and sold does not mean it has patentable utility.

With respect to credibility, applicant is reminded that in order to meet the requirements of 35 USC 101, the specification must disclose at least one utility that is specific and substantial, as well as credible (absent a showing of well established utility, which would presume that the utility was credible). The claims have been rejected because 1) the specification fails to disclose at least one utility that is both specific and substantial, and 2) no convincing evidence has been presented to show that an EST, for

which only its nucleotide sequence and source have been disclosed, has a well established utility.

The brief does not appear to directly argue for a well established utility for the claimed invention; however, the arguments concerning the commercial value of ESTs in general (brief, pages 12-13) may implicitly be directed to a well established utility for any EST in general, and the claimed nucleic acid molecules in particular. However, such evidence is not relevant to 35 USC 101.

The examiner maintains that the uses asserted for the claimed invention are methods where the claimed invention is, itself, an object of scientific study, e.g. to determine whether the corresponding genomic DNA of maize or corn contains a polymorphism that can be detected with the claimed invention. The specification cannot enable or tell how to use the invention within 35 U.S.C. 112, first paragraph, if there is no patentable utility within 35 U.S.C. 101. The examiner maintains that there is no patentable utility for the claimed invention for the reasons set forth above and thus the claims are not enabled.

Insofar as the specification fails to describe a specific, substantial, and credible utility for SEQ ID NO: 1 itself, so does it fail to describe a specific and substantial utility for the nucleic acid molecules that hybridize to SEQ ID NO: 1 (see for example claim 1), are complementary to SEQ ID NO: 1 (see for example claim 1), contain additional elements (see for example claims 2-3 and 5-7), are fragments of SEQ ID NO: 1 (see for example claim 21), or have a level of similarity with SEQ ID NO: 1 (see for example claim 23).

Claims 1-7 and 20-24 are also rejected under 35 U.S.C. 112, first paragraph. Specifically, since the claimed invention is not supported by either a specific, substantial, and credible asserted utility or a well established utility for the reasons set forth above, one skilled in the art clearly would not know how to use the claimed invention.

The uses asserted for the claimed invention are methods where the claimed invention is, itself, an object of scientific study. The specification cannot enable or tell how to use the invention within 35 U.S.C. 112, first paragraph, if there is no patentable utility within 35 U.S.C. 101. The examiner maintains that there is no patentable utility for the claimed invention for the reasons set forth above and thus the claims are not enabled.

Claim Rejections - 35 USC § 102

Claims 1-7, and 22-24 are rejected under 35 U.S.C. 102(a) as being anticipated by Tikhonov et al. (PNAS, 96:7409-7414, 22 June 1999) in view of GenBank Accession No. AF123535.

Tikhonov et al. discloses sequencing and analyzing regions of the maize genome. Bacterial articificial chromosomes and *E. coli* were used. GenBank Accession No. AF123535 is referenced for all sequences. (See at least abstract; page 7409, section on materials; Figure 1; and page 7411, section on analysis of the Maize region.)

NCBI Accession No. AF12535 discloses a nucleotide sequence from *Zea mays*. In particular, nucleotides 57344-57428 have significant similarity to nucleotides 196-280 of SEQ ID NO: 1. See attached alignment. For example, the 14 nucleotides GAG TTC CTC GGC TC match exactly. The 42 nucleotides CCG GAG GCG TAA GAG TTC CTC GGC TCG GTC GGG CTT GCC CCT have complementarity with five

mismatches. If the previously identified 14mer subsequence of SEQ ID NO: 1 is chosen as the second nucleic acid molecule having a sequence of SEQ ID NO: 1 or a complement thereof," this 14mer sequence will hybridize to the sequence of AF123535 under the conditions named. If the previously identified 42mer subsequence of SEQ ID NO: 1 is chosen as the second nucleic acid molecule having a sequence of SEQ ID NO: 1 or a complement thereof," this 42mer sequence will hybridize to the sequence of AF123535 under the conditions named, even assuming a 1.5 degree drop in melting point for each percentage of mismatch. This assumption is standard in the art. See attached oligonucleotide properties calculator. Note that the use of the article "a" includes subsequences of SEQ ID NO: 1. Applicant has not disputed that this language includes subsequences of SEQ ID NO: 1 nor have they amended the claims to make clear that hybridization to the entirety of SEQ ID NO: 1 is intended such as by a recitation of "a second nucleic acid molecule having **the** sequence of SEQ ID NO: 1" or "a second nucleic acid molecule having the complete sequence of SEQ ID NO: 1." With respect to claims 2, 3, and 6-7, Tikhonov et al. makes clear that polymorphisms, promoters, and microsatellites are present in the sequence of AF123535. (Again, see at least Figure 1.) With respect to claims 22-24, the language "a nucleic acid sequence" embraces a subsequence of SEQ ID NO: 1 and the sequence of AF123535 has 100% identity with at least the particular 14mer subsequence identified above.

Applicant is reminded that the claimed invention is not directed to SEQ ID NO: 1 but rather sequences that hybridize to or have particular identity with parts of SEQ ID NO: 1.

Should applicant traverse this rejection, they are requested to provide their calculation or empirical determination of melting temperature with all assumptions as well as basis for their interpretation of what the claims require should it differ from that set forth by the examiner above.

Conclusion

No claim is allowed.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Marianne P. Allen whose telephone number is 571-272-0712. The examiner can normally be reached on Monday-Thursday, 5:30 am - 1:30 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Michael Woodward can be reached on 571-272-0722. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Marianne P. Allen
Primary Examiner
Art Unit 1631

mpa

MICHAEL P. WOODWARD
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 1600

2/13/04

Notice of References Cited	Application/Control No. 09/606,977	Applicant(s)/Patent Under Reexamination BYRUM, JOSEPH R.	
	Examiner Marianne P. Allen	Art Unit 1631	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Tikhonov et al., PNAS, 96:7409-7414, 22 June 1999.
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

June 22, 1999

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 7409–7414, June 1999
Genetics

Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum

ALEXANDER P. TIKHONOV, PHILLIP J. SANMIGUEL, YUKO NAKAJIMA, NINA M. GORENSTEIN, JEFFREY L. BENNETZEN, AND ZOYA AVRAMOVA*

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392

Communicated by John D. Axtell, Purdue University, West Lafayette, IN, April 29, 1999 (received for review October 12, 1998)

ABSTRACT **Orthologous *adh* regions of the sorghum and maize genomes were sequenced and analyzed. Nine known or candidate genes, including *adh1*, were found in a 225-kilobase (kb) maize sequence. In a 78-kb space of sorghum, the nine homologues of the maize genes were identified in a colinear order, plus five additional genes. The major fraction of DNA in maize, occupying 166 kb (74%), is represented by 22 long terminal repeat (LTR) retrotransposons. About 6% of the sequence belongs to 33 miniature inverted-repeat transposable elements (MITEs), remnants of DNA transposons, 4 simple sequence repeats, and low-copy-number DNAs of unknown origin. In contrast, no LTR retroelements were detected in the orthologous sorghum region. The unconserved sorghum DNA is composed of 20 putative MITEs, transposon-like elements, 5 simple sequence repeats, and low-copy-number DNAs of unknown origin. No MITEs were discovered in the 166 kb of DNA occupied by the maize LTR retrotransposons. In both species, MITEs were found in the space between genes and inside introns, indicating specific insertion and/or retention for these elements. Two adjacent sorghum genes, including one gene missing in maize, had colinear homologues on *Arabidopsis* chromosome IV, suggesting two rearrangements in the sorghum and three in the maize genome in comparison to a four-gene region of *Arabidopsis*. Hence, multiple small rearrangements may be present even in largely colinear genomic regions. These studies revealed a much higher degree of diversity at a microstructural level than predicted by genetic mapping studies for closely related grass species, as well as for comparisons of monocots and dicots.**

The grasses belong to a family of monocotyledonous angiosperms that are well differentiated morphologically from the other angiosperm families and have a single (monophyletic) origin. Their genome sizes, however, may vary a great deal between species. Thus, rice has an estimated genome size of 430 megabases, which is ≈11× smaller than barley, 6× smaller than maize, and 2× smaller than sorghum. These large differences in genome sizes, coupled with differences in the degree and the nature of their investigations, have obscured some common features of grass genomic design. Recent studies comparing high-density linkage maps with DNA markers revealed extensive synteny of chromosomal segments between related species (1–5). Valuable as it is, full genome recombinational mapping of DNA markers is not an efficient approach for detecting small rearrangements. Because the available high-resolution maps based on completed nucleotide sequence are largely restricted to individual genes and their proximal neighborhoods, we are left with two obvious questions that cannot be answered at a full-genome level of analysis. These questions are, will the colinearity observed at the 2- to 20-centimorgan level, the sensitivity level of standard recombinational mapping, be preserved or will it break down at a local level (5), and what will the pattern of gene distribution be, relative to the noncoding, nongene-containing portions of the chromosomes? Currently, two scenarios are considered for how genes could be distributed in complex grass genomes (6): a scattered distribution of genes among noncoding regions, leaving large distances between neighboring genes, and a clustered organization of genes, segregating large regions of transcribed DNA from surrounding large blocks of noncoding DNA. Gross analysis of both animal and plant genomes, based on density gradient centrifugation (isochore analysis), on DNA gel blot hybridization and *in situ* hybridization, support the second model: homogeneous, gene-enriched islands, >100 kilobases (kb) in size, segregated amidst a sea of repetitive DNAs (7–9). However, results from sequencing projects and studies at a molecular level have provided data on large stretches of genomic sequence, which may challenge our traditional views of genome organization. Thus, in *Drosophila*, functional genes were found not only in euchromatin, but also in the heterochromatic pericentric regions (reviewed in ref. 10), arguing that overall genome analysis cannot provide an adequate picture of genome microstructure.

Until very recently, the largest uninterrupted sequence data available from a complex grass genome constituted ≈60 kb from barley chromosome 4HL (11). It revealed the presence of three genes and one retrotransposon. However, the lack of data at a similarly detailed level from a syntenous region of a related species inevitably limited the scope of possible conclusions, particularly those concerning its evolution. The power of comparative genome analysis at a microcolinear level for gene identification and for characterization of intergene spaces in maize, sorghum, and rice has already been demonstrated (12–15). Here, we focus on a comparative analysis of orthologous *adh* regions in maize (*Zea mays*) and sorghum (*Sorghum bicolor*). Sorghum is a close relative of maize, with an ≈3.5-fold smaller genome (16). In this study, we demonstrate coexistence, in a 225-kb region of maize chromosome 1, of a mixture of gene clusters and individual genes interspersed with 14- to 70-kb blocks of highly repetitive DNA. In addition, we compared the sequences from the two grass genomes to homologous *Arabidopsis* regions available in the databases. The molecular characterization and comparison of the three plant genomes allowed us to draw a clearer picture of the structure and evolution of these regions.

MATERIALS AND METHODS

Materials. Overlapping yeast artificial chromosomes (YACs) 334B7 and 119E3, carrying a maize insert from the *adh1* region (17), were used. A maize bacterial artificial chromosome (BAC) (86A10) originating from maize line B73 was obtained from Genome Systems (St. Louis). Sorghum BAC 110K5 was obtained from the BAC library previously described (18). *Escherichia coli* strains DH5, DH10B, LE392, KW251 (Promega), and XL1 Blue (Stratagene) were used. Plasmid DNA for cloning and sequenc-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: LTR, long terminal repeat; YAC, yeast artificial chromosome; BAC, bacterial artificial chromosome; EST, expressed sequence tag; SSR, simple sequence repeat; MITE, miniature inverted-repeat transposable element.

*To whom reprint requests should be addressed. e-mail: zavramov@bilbo.bio.purdue.edu.

ing was prepared according to Del Sal *et al.* (19). λ phage DNA was prepared according to Sambrook *et al.* (20).

DNA Manipulation and Gel Blot Analysis. All standard recombinant DNA procedures were conducted according to enzyme manufacturer's recommendations or as described (19, 20). Reverse Southern hybridizations were performed by digesting BAC or YAC subclones with alternative sets of restriction enzymes to generate overlapping fragments. Gels were blotted onto nylon filters and were blotted and probed with radioactively labeled total genomic DNA of sorghum (21).

Nucleotide Sequencing and Computer Analysis. DNA sequencing was conducted by the *Tn*1000 strategy (22), modified by using a rifampicin resistant derivative of *E. coli* strain DH5 as a recipient and XL1 Blue as a donor strain. Programs used for sequence analysis were the GCG package v.8 and v.9 (23), BLAST (24), AAT at http://genome.cs.mtu.edu/aat/aatdoc.html (25), and REPEATMASKER (http://ftp.genome.washington.edu/RM/RepeatMasker.html).

The nucleotide sequences were intensively analyzed for the presence of direct and/or inverted repeats, retroelements, and/or other repetitive DNA by six different programs: BLAST, COMPARE, FASTA, REPEAT, STEMLOOP, and REPEATMASKER. Direct comparison of maize and sorghum sequences and sorghum and *Arabidopsis* sequences was conducted by using the program COMPARE. Graphical outputs of the comparisons were obtained by the DOTPLOT program. Insertions/deletions were identified by the GAP program, with gap weight 100 and length weight 0.

RESULTS

The Sorghum Region: Composition and Organization

A portion of sorghum BAC 110K5, containing an ortholog of the maize *adh1* gene, was sequenced. A contiguous 78-kb nucleotide sequence was generated with an average redundancy of 5.8 (GenBank accession no. AF124045). The search for genes was carried out by applying four criteria: (*i*) comparison to GenBank databases; (*ii*) comparison to cDNA sequences generated in the lab or expressed sequence tag (EST) databases; (*iii*) a search for exon/intron structure, as defined by the internet-based gene prediction programs and by a comparison (gapped alignment) to ESTs; and (*iv*) homology to sequences in the orthologous maize and *Arabidopsis* regions.

Fourteen gene candidates, including the *adh* gene, were identified. They were assigned numeric names in the order of their appearance on the BAC (Fig. 1). Only one gene, 110K5.3, corresponds to a gene with a known function, the pollen specific *adh* gene, homologous to maize *adh1*. Six gene candidates exhibited varying degrees of homology to known or putative proteins. These genes are 110K5.1, similar to *Mesocricetus auratus* guanine nucleotide-binding protein ($3.1e^{-36}$, U13152); 110K5.6, homologous to *Arabidopsis thaliana* carbohydrate kinase ($3e^{-07}$, 91N13T7); 110K5.7, homologous to the human cyclinH ($2e^{-12}$, P51946); 110K5.8, homologous to the yeast small GTP-binding protein ($4e^{-42}$, 010190); 110K5.9, homologous to the small nuclear ribonucleoprotein Prp4p of *A. thaliana* ($2e^{-26}$, 022212YG62); and 110K5.14, homologous to various plant chloroplast ATPase synthase Δ subunits: tobacco ($9.1e^{-12}$, P32980), pea ($9.2e^{-10}$, 002758), and spinach ($2.5e^{-11}$, 170098).

The remaining eight genes were predicted on the basis of homology to various plant ESTs and cDNAs. Thus, comparison to maize cDNAs (AF124736, AF124737, AF124738, AF124739, AF124740) outlined putative exons for four gene candidates: 110K5.1, 110K5.2, 110K5.4, and 110K5.8. Two adjacent maize ESTs (AA979993 and AA979792) with high homologies ($2.6e^{-109}$ and $1e^{-64}$, respectively), suggested a poly(A) tail region and putative exons for 110K5.12 gene. An ORF of 1,027 bp was predicted. An observation, important for our conclusions later, was the homology found between two maize cDNAs (AF124737, AF124739) and candidate gene 110K5.8 because the respective maize homologue is one of the genes missing from the colinear maize contig. Putative exons for two more gene-candidates (110K5.5 and 110K5.4) were outlined based on homologies found to *A. thaliana* sequences from BAC T10M13 (ATAF001308).

A rice cDNA (DDBJ RICS1659a) is homologous to three putative terminal exons of 110K5.11 gene, designated *u22* in an earlier study (12). Another rice cDNA (28880) displays some homology ($4e^{-20}$) over a 200-bp region of what could be a sorghum gene 110K5.13. The homologous maize region is missing, and it is difficult to recognize other features of a potential gene. This sequence, however, might contain a separate gene

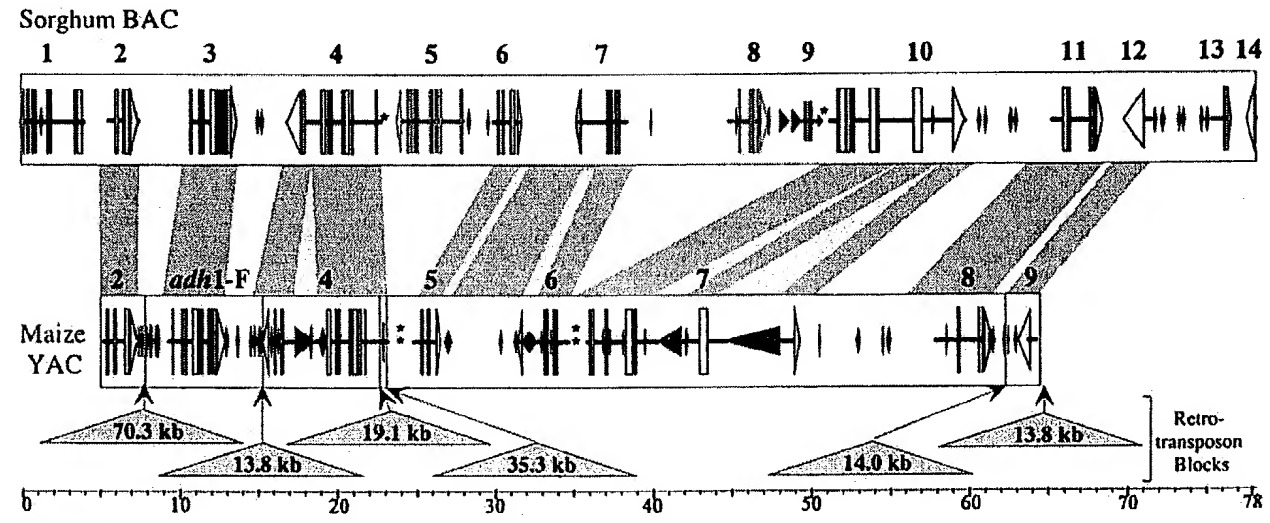


FIG. 1. Schematic representation of the structures of orthologous *adh* regions in maize and sorghum. The upper bar represents sorghum BAC 110K5. Putative genes are numbered by the order of their appearance on the contig, and the putative exons/introns are shown by boxes and connecting lines, respectively. The open triangles mark the end and the direction of transcription for the established cases. The red diamonds show the location and the size of MITEs and other transposon-like elements. The genes shown in blue are genes, lying among orthologous genes but missing from the maize contig. The stars reflect the location of simple repetitive DNAs. The composition of the maize region, derived from YAC 334B7, is shown below. Putative genes, exon/introns, and inserted small elements are marked as above. The purple triangles show the location, the size, and the orientation of transcription of the putative LINEs. The triangles show the location and sites of insertion of the retroelement blocks, and the numbers indicate the block sizes. The shaded regions connecting the maize and the sorghum contigs outline the regions of sequence homology between the two species. The lighter strips within correspond to stretches of interrupted homology, usually associated with insertion of small mobile elements. Both contigs are shown in scale.

because it would have a direction of transcription opposite to both flanking 110K5.12 and 110K5.14 genes.

Analysis of reiteration frequencies of the various classes of DNA on the sorghum *adh* BAC (determined by reverse Southern hybridization) suggested that the region was composed mainly of low-copy-number and some middle repetitive DNAs (data not shown). The nucleotide sequences were analyzed for the presence of direct and/or inverted repeats, retroelements, and other repetitive DNA by six different programs (see *Materials and Methods*).

No obvious long terminal repeat (LTR)-retroelements were revealed. However, numerous transposon-like elements and simple sequence repeats (SSRs) were identified (Fig. 1). Seventeen of the putative transposon-like elements were found in the intergenic space, and three were found in introns. For 12 of them, homologies to the *Tourist* elements found in the *sh2-a1* region of sorghum (15) were revealed. Eleven of them are in the spacer regions between genes, and one is inside an intron of the 110K5.8 gene. One *Tourist* element is found inserted within another *Tourist*, in the spacer between *adh* and the 110K5.4 gene. Two direct moderately repetitive sequences, 516 bp and 634 bp long, displaying >70% homology to each other, were found between 110K5.8 and 110K5.9 genes. SSRs are represented by two $(AT)_{25}$ and $(AT)_{67}$ stretches and by three islands of $(GGA)_{46}$, $(CGG)_{30}$, and $(GAA)_{26}$ repeats. In summary, ≈66 kb (85%) of the 78.2 kb of contiguous sorghum DNA around the pollen *adh* gene represent the gene-containing fraction, containing 14 genes. The remaining 12 kb (15%) belong to miniature inverted-repeat transposable elements (MITEs), SSRs, and DNA lacking an identified origin.

Analysis of the Maize Region

A contiguous series of restriction fragments (contig), covering 258 kb from YAC 334B7, containing the maize *adh1*-F gene (26), has been the focus of our studies. Previous work has established the overall structure of the contig, the location of the only known gene (*adh1*-F), the spatial segregation of the low-copy-number regions from highly reiterated DNA, and the nature and organization of the dispersed repetitive DNAs (26–28). Studies based on a cross-referencing approach, using cross-hybridization with the orthologous sorghum region, identified *adh1* and several other potential genic regions (12). To determine the nature and the microstructure of these possible genic regions, all low-copy-number regions were sequenced with a higher redundancy whereas regions consisting of retroelements of repetitive nature were sequenced with lower redundancy (average redundancy of 4.8). The available sequence data, therefore, represent 217.9 kb of maize chromosome 1, spanning a region of ≈225 kb. The only gaps in the sequence, corresponding to a total of 7.4 kb, were located within highly repetitive retroelements (in the internal part of *Grande*-zm1, *Ji*-1, and *Ji*-6 and inside *Kake*-1) (Fig. 3). The sequences of some retrotransposon LTRs in this region were reported previously (27, 29). All sequences from the maize contig have GenBank accession no. AF123535.

Structure of the Low-Copy-Number Gene-Containing Fraction

Nine putative genes were identified and designated in the order of their appearance on the YAC. Gene 334B7.1, located at the left end of the YAC, has been sequenced only partially and will not be discussed further. 334B7.2 is the first gene at the 5′ end of the contig, located 70.3 kb upstream of *adh1*. It is homologous to the sorghum gene 110K5.2 from the colinear region. A typical gene structure was established for 334B7.2: a putative TATA box and 5 exons, four of which are homologous to cloned cDNA (AF124736). There is a 700-bp ORF in the same orientation as *adh1*, and the respective homologous regions in the two species are of similar length, ≈2.4 kb.

334B7.3. In the numeric system adopted in this study, this corresponds to the *adh1* gene. It is homologous over a region of 3.8 kb to the sorghum region 110K5.3, which contains the *adh1* ortholog.

334B7.4. The closest downstream neighbor to *adh1* is separated by a 13.8-kb block of retrotransposons. Comparison to the maize cDNA (AF124740), to ESTs from various sources (maize, *A. thaliana*, *Caenorhabditis elegans*, human, W49897, N65123, 2104528, 2088768, HSPD06991, respectively), and to the colinear sorghum and *Arabidopsis* (ATAF001308) sequences revealed eight putative exons.

A very small region, 421 bp of low-copy-number DNA, is located 19.1 kb downstream, separated by a block of repetitive DNA. It is interesting to note that, in a previous study, by using cross-hybridization as a tool for gene identification, we were able to locate this small, low-copy-number fragment embedded in a sea of highly repetitive DNA (12). Although the nature and the origin of this sequence was completely obscure at that time, its high level of conservation between maize and sorghum suggested that it might bear some function (12). Here, analysis at a sequence level and comparison to the homologous sorghum 110K5.4 gene helped us recognize it as a region containing the putative first exon of the 334B7.4 gene. In maize, it has been separated from the rest of the gene by the insertion of two unrelated retroelements, *Huck*-2 and *Fourf*, displacing it by ≈19 kb. A third retroelement, *Milt*, has inserted into the 3′ untranslated end of the gene, followed by a more recent insertion of *Opie*-2 into *Milt* (29). As a result, the genomic space occupied by portions of this gene is scattered over 42 kb. The eight predicted exons have uninterrupted ORFs with the opposite orientation from *adh1*, but it remains to be tested whether this gene is expressed. A 1.13-kb LINE was found within a putative intron.

334B7.5, 334B7.6, 334B7.7, and 334B7.8. Proceeding further to the right (Fig. 1), after a 35.3-kb block of highly repetitive DNA, comes a large region, 39.2 kb, of low-copy-number DNA. This is the largest segment in the maize genome, observed by us so far, that lacks a detected LTR-retrotransposon. Analysis of its sequence suggested the presence of at least four genes. Because no cDNAs or ESTs were available to identify putative transcription starts or stop codons for 334B7.5 and 334B7.6, the first two genes in the group, the boundaries of their space were defined by the beginning and by the end of homology to the corresponding sorghum region. As shown by the shaded areas between the two contigs (Fig. 1), the homology stretches for ≈10 kb. Comparison at the predicted amino acid level found homology for 334B7.5 to *A. thaliana* carbohydrate kinase cDNA, as found and described above for the sorghum 110K5.6 gene.

Homology to sorghum (110K5.7 gene) and to the human cyclin H protein ($2e^{-12}$, P51946), determined the position of the putative maize gene 334B7.6. It is interesting to point out that, between the regions occupied by these putative genes, there is ≈5 kb of sequence homology between sorghum and maize, interrupted only by inserted elements. At this stage, we cannot decide what the region between the two putative genes contains, but its conservation indicates that it might have biological relevance. Whether it contains regions belonging to the already identified genes, or a different gene, remains to be established.

Surprisingly, there are only 200 bp of unconserved sequence separating 334B7.6 from 334B7.7, apparent genes, to be transcribed in opposite directions. This fact was unexpected because it provides a first example of two closely positioned but otherwise unrelated genes in maize. In contrast, the respective sorghum homologous genes, 110K5.7 and 110K5.10, are 12 kb away from each other. The nature of this intervening sorghum DNA will be discussed in detail below.

The space occupied by the maize gene-candidate, 334B7.7, spans ≈15.1 kb. Its homology with the sorghum 110K5.10 gene is interrupted only at the sites of inserted elements, two of which are LINEs (related to *Colonist1* and *Colonist2* elements, ZMU90128). A putative TATA box and ATG codon were

located. Five putative exons were predicted by BLASTX, and gapped alignment of sequences with homology to the putative gene 11 on *A. thaliana* BAC T5I7 (2642163). Comparative analysis with several cDNAs isolated from a maize seedlings library provided further evidence that this region encompasses a gene (data not shown).

The last of the predicted genes in this 39-kb low-copy-number DNA stretch, 334B7.8, is 6.1 kb downstream of 334B7.7, occupying 4.8 kb of genomic space. Earlier, it was predicted that this region might contain a gene because a homologous rice EST (DDBJ RICS1659a) was found (12). Here, we managed to identify the putative TATA-box, a start codon, a poly(A) signal, four putative exons, and an ORF in the same orientation as *adh*.

334B7.9. About 14 kb downstream of 334B7.8 is the site of the last predicted gene in the sequenced maize region. It is homologous to the sorghum 110K5.12 sequence and to two maize cDNAs (AA979993 and AA979792) with 96 and 93% homology, respectively.

Organization of the Repetitive DNA on the Maize Contig

Earlier, a model for the organization of the repetitive DNA in the maize *adh1* region was proposed based on the order of 37 classes of repeats of the contig and the diagnostic sequencing of putative LTRs (27). Nineteen nested LTR retroelements and two solo LTRs were identified (27). Subsequently, as a result of complete sequencing of most retroelement blocks, regions of ambiguity have been clarified: a retroelement from the *Cinful* family, *Cinful*-2, and an older element, *Tekay*, have been located immediately upstream of *adh1* (29). *Rle*, an even older retroelement, has provided an insertion site for ≈70 kb of retrotransposons, occupying a block at the 5′ end of the contig and 3′ the 334b7.2 gene (Fig. 3).

Overall, the LTR-retroelements comprise 166.4 kb (74%) of the available sequence. The high- and middle-repetitive retroelements are arranged in six blocks of different sizes on the contig: 70.3 kb, 13.8 kb, 19.1 kb, 35.3 kb, 14.0 kb, and 13.9 kb (Fig. 1). The first block of 70.3 kb (with internal gaps in our sequence data of 6.8 kb) covers all of the intergenic space between 334b7.2 and *adh1* genes. The next two blocks are found inside the space occupied by the 334b7.4 gene, and the 35.3-kb block separates the 334b7.4 and 334b7.5 genes. The fifth block of retroelements, containing nested *Reina* and *Cinful*-1 retrotransposons, is inserted between two genes (334B7.8 and 334B7.9), placing them 14 kb apart. The last retroelement block contains nested *Kake*-1 and *Kake*-2 retrotransposons, covering a region of 13.9 kb.

The whole region has been analyzed for the presence and the distribution of mobile DNAs and simple sequence repeats. Because of the overwhelming amount of data and the complexity of such an analysis, at this stage, MITEs were identified by homology to already existing MITE sequences in the databases or by the presence of DNA sequences flanked by defined repeats and host insertion duplications (30, 31). Most of these putative MITEs were found within low-copy-number sequences, as short inserts interrupting the homology between maize and sorghum. Not a single MITE was identified within the space occupied by the retroelements.

A total of 33 DNA transposon-like elements were recognized to this end: For 8, homology to the *Tourist* elements was discovered; for 3, homology to the *Castaway* family was discovered (30). Two *Tourist* elements and a *Tourist/Castaway* inserted into each other are located 5′ of the *adh1* gene. A set of three inserted MITEs is found 3′ of the 334B7.8 gene.

Immediately 5′ to the 334B7.2 gene is a 50-bp sequence with 86% identity to the *Sleepy* transposon of maize (ZMU28041), followed by two closely positioned, almost perfect, 44/46 bp, direct repeats. A 230-bp remnant of a maize *Ds* insertion element (X51632) is found between the 334B7.7 and 334B7.8 genes.

Finally, ≈200 bp of simple repeats (CGG and TGG) are found in the space 5′ to the 334B7.5 gene; several (CT)n clusters and a 28-bp stretch of uninterrupted Cs are at the immediate 5′ end of 334B7.7, and a 62 bp-long SSR is found between 334B7.7 and 334B7.8. In summary, the 225.5 kb of the maize sequence around the *adh1* gene is composed of ≈74% LTR-retroelements and 20% genic DNA. The remaining 6% is made up of MITEs, insertion elements, SSRs, and DNA lacking identified origin.

Comparison of the *adh* Regions in Maize and Sorghum

Nature of the Unconserved Sorghum Sequences. Direct comparison of the two colinear regions revealed a patchy pattern of homology (Fig. 2). The diagonal lines, indicating regions of homology, correspond to the gene-coding sequences and their immediate flanking regions whereas the gaps between the diagonal lines represent unconserved sequences. Examining the plot, it is seen that unconserved regions belong to two major categories: missing genic sequences and intergenic spacers.

Three of the fourteen putative genes in sorghum, 110K5.1, 110K5.13, and 110K5.14, are beyond the sequenced regions of the maize contig analyzed here. Three other genes, 110K5.5, 110K5.8, and 110K5.9, were not found in the maize region, although they lie among orthologous genes (Fig. 2). The possibility that the maize homologues were artifactually deleted during the isolation of YAC 334B7 was tested. A second YAC, 119E3, and maize BAC 86A10, covering the controversial area, were hybridized to a probe homologous to the maize 334B7.5 gene, as a positive control, and to two probes recognizing 110K5.5 and 110K5.8 genes from the 4-kb and 12-kb regions missing in maize, respectively. The results indicated that maize homologues of the sorghum genes were not present on either of the new clones (data not shown). These genes, however, are present elsewhere in the genome because they hybridized to the sorghum probes in gel blot hybridizations to total maize genomic DNA (data not shown).

Because the expected loci for both missing genes would have been flanking the maize 334B7.5 and 334B7.6 genes, special attention was devoted to the DNA flanking them in maize and the respective DNA in sorghum. The 1.4 kb between 110K5.4 and 110K5.5 in sorghum displayed no obvious features other than a 20-bp stretch of alternating (A)n and (T)n. In the respective maize region, the DNA between the two genes (over the gap of the missing homologue of 110K5.5) is a 2.1-kb sequence rich in (CGG)n and (TGG)n repeats. A BLAST search found homologies to human fragile DNAs (AF012603 and U48436).

The other two tightly linked sorghum genes, 110K5.8 and 110K5.9, whose homologues are absent from the maize region, are separated from a neighboring (110K5.7) gene by 6 kb of DNA for which no unusual features were recognized. At the other end, in 0.6 kb of spacer DNA separating 110K5.9 from 110K5.10, an

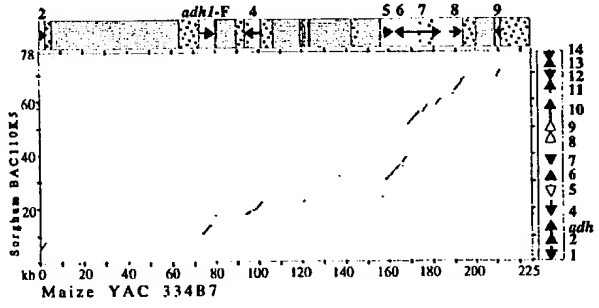


FIG. 2. DOTPLOT homology comparisons for the colinear maize and sorghum regions. The location of the high- and middle-copy-number retroelement blocks is shown by the shaded and dotted boxes, respectively. The open boxes with the arrows show the putative maize genes identified on the contig. On the vertical line, the sorghum BAC is shown. The diagonals, reflecting homologous regions, coincide with the regions taken by genes. The sorghum genes 110K5.5, 110K5.8, and 110K5.9 missing from maize are shown by lighter arrows on the sorghum BAC.

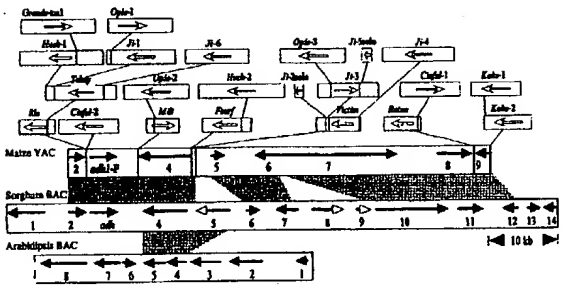


FIG. 3. A simplified schematic comparison between the contigs of maize (YAC 334B7), sorghum (BAC 110K5), and *Arabidopsis* (BAC T10M13). The arrows show the location and the orientation of the putative genes. The shaded regions outline the regions of colinearity between the three species.

SSR with $(CGG)_{10}$ repeats was found. In maize, a short 200-bp spacer separates the neighboring 334B7.6 and 334B7.7 genes. This linking DNA contains two $(CT)_{5-9}$ bracketing a 28-bp stretch of uninterrupted Cs.

In summary, 66.4 kb of sorghum sequence is colinear to 211.7 kb of maize in the orthologous *adh* regions. The overall amount of conserved DNA between the two species is 22% for maize and 57% for sorghum whereas the unconserved fractions represent 78% and 43% of the maize and sorghum DNAs, respectively. The majority (74%) of the maize DNA is composed of LTR-retroelements in the intergenic spaces whereas not a single LTR-retrotransposon was detected in the sorghum region.

Comparisons to the *Arabidopsis* Genome

The missing genes in maize raise a question of whether they have been deleted from the colinear maize chromosome or whether they inserted into the sorghum genome at their present position after the two species diverged, some 15–20 million years ago (32). A possible answer for one of the genes, 110K5.5, is suggested by the finding that two genes, homologous to adjacent sorghum genes, 110K5.4 and 110K5.5, were found next to each other in *Arabidopsis* BAC T10M13 (AF001308) (Fig. 3). Their homology to the respective *Arabidopsis* T10M13.5 and T10M13.4 genes, their tight linkage, and their conserved orientation all suggest that these genes were linked in the ancestor of monocots and dicots. This fact suggests that the maize homologue of 110K5.5 was deleted from the *adh1* region after the separation of maize and sorghum.

All sequence data generated from this study were compared with the sequence information available for *Arabidopsis*. No other colinear regions were identified, although syntenous genes, homologous to the sorghum 110K5.8 and 110K5.10 sequences, were found on *Arabidopsis* chromosome II (AC002510 and AC003000). In the case of *Arabidopsis*, however, the homologous sequences are on two nonoverlapping BACs, separated by at least 100 kb.

DISCUSSION

The importance of small reference genomes and the power of comparative genomics for gene discovery and for evolutionary studies have been demonstrated for bacteria (34, 35), animals (36, 37), and plants (5, 38, 39). The obvious advantages of the smaller and simpler sorghum genome, the fact that differences in genome size do not correlate with morphological and physiological complexity of the organisms (16, 21, 40), as well as the observed similarity and colinearity between the maize and sorghum genomes (1, 12, 41) suggested sorghum as an important model organism for the characterization of the maize genome. Subsequently, the high level of gene homology and synteny found between rice and sorghum (13–15), rice and wheat (38), and rice and barley (42), coupled with the small size of the rice genome, provoked the idea that description of the rice genome alone will indicate the gene content and the key aspects of genome organization for all grasses (42).

The results of this study provide an example of the extent to which such assumptions may be correct. Despite a 3.5-fold difference in total genome size (2,500 megabases for maize and 750 megabases for sorghum) (16), the two grass genomes have a similar set of unique sequences (1, 41) and a general colinearity that, at least in the orthologous genomic regions of the *adh* loci, is well conserved (12).

The major unconserved fraction belongs to the intergenic DNA that, in maize, is represented by 22 nested LTR-retroelements, 33 MITEs, SSRs, and DNA of an unidentified origin. A remarkable distinction between these genomes is that no LTR-retrotransposons was found in the 78-kb continuous sequence in sorghum.

A notable feature of the distribution of MITEs is that they are found within introns or in proximal regions flanking genes, as observed earlier for various grass genes (reviewed in ref. 30). Our data demonstrate that, within 166 kb of DNA composed of LTR-retrotransposons, not a single MITE was found. Hence, MITEs must be unable to insert and/or be retained in these methylated, presumably heterochromatic regions (43). Alternatively, the MITEs may have arrived before the appearance of the retroelements, ≈2–6 million years ago (29). A gene-specific insertion preference of MITEs is similar to that of other maize inverted-repeat transposable elements, like *Mutator* (44).

It was speculated that, after "subtracting" the unconserved fraction, exact colinearity of the gene-containing portions of the two genomes might be observed. However, three genes were found to be deleted from the maize continuum. The true absence of these genes from the maize chromosomal *adh1*-F region was confirmed by testing an additional maize YAC and a maize BAC that cover the investigated area. Homologues of these sorghum genes are present in the maize genome, however, because they hybridized to the sorghum probes in gel hybridizations to maize genomic DNA. In addition, two maize cDNAs were found homologous to the sorghum 110K5.8 gene, supporting a conclusion that the respective maize gene(s) are present, and active, elsewhere in the maize genome.

These results suggest that, even in largely colinear genomic regions, multiple small rearrangements may be present. In maize, such events might be tolerated because of its tetraploid nature (45, 46) and may reflect the "fluidity" of the maize genome caused by the retroelements (47) or other mobile DNAs. Therefore, despite the general principle of similarity and colinearity of the grass genomes, sequence analysis at a microstructural level may reveal significant divergence. For instance, we have found a complete lack of colinearity between the region carrying the *adh1* orthologous locus of rice to the *adh1* region of maize (Y.N., P.J.S., A.P.T., Z.A., H. Zhang, R. Wing, and J.L.B., unpublished work). Although this result provides valuable information about the evolution of the region in parallel with speciation, it also illustrates the risk of making general conclusions based solely on overall genome analysis.

Although a two-gene colinearity was found between sorghum genes 110K5.4–110K5.5 and *Arabidopsis* genes T10M13.5-T10M13.4, the two flanking *Arabidopsis* genes, T10M13.3 and T10M13.6, are not linked to each other in either maize or sorghum. This indicates a minimum of two chromosomal rearrangements in a stretch of four genes. Two other tightly linked sorghum genes, 110K5.8 and 110K5.10, have homologues located on *Arabidopsis* chromosome II. In the case of *Arabidopsis*, however, the genes are not tightly linked because they were discovered on different BACs, >100 kb apart. This result indicates that, although occasional two-gene, or somewhat larger, conserved linkages will be observed between monocots and dicots, long conserved segments may be rare. This is a surprising result. If, as it appears, long sequence blocks approaching whole chromosome or chromosome arms have been conserved in the 50–100 million

years since the grasses diverged from each other, then why have there been so many rearrangements in the approximately 200 million years since monocots and dicots diverged? Because intradicot genome colinearity appears to be less than that for the grasses, it is possible that most of these rearrangements occurred early and often in the dicot descendants of the primordial angiosperm. Alternatively, the lineage that has given rise to *Arabidopsis* may have undergone an exceptionally large number of rearrangements in recent evolutionary time.

A major observation of our studies is the pattern of gene arrangement in the related sorghum and maize genomes. Thus, nine genes in a 225-kb region provide an average gene density of about one gene per 25 kb in maize. In sorghum, the density we observe is one gene per 5.6 kb. Assuming that there are ≈30,000 genes in sorghum (genome size of 750 megabase pairs) and ≈50,000 genes in maize (genome size of 2,500 megabase pairs), the expected average gene density would be one gene per 25 kb in sorghum and one gene per 50 kb of maize. The total gene numbers here are crude, of course, but are based on a predicted 25,000 or so genes for *Arabidopsis* (33). Sorghum maps as a diploid with a reasonable amount of segmental duplication (1, 3) whereas maize is an ancient tetraploid (32), suggesting respective gene numbers of 30,000 and 50,000. Therefore, the experimentally determined distances indicate a much denser packing of genes than expected, especially for the sorghum region. In sorghum, genes are evenly distributed along the 78 kb studied. A similar pattern was observed earlier for the genes in the *sh2-a1* region (13–15). This pattern also would fit into a gene-cluster model because these regions might actually represent isolated, gene-enriched segments of sorghum chromosomes. Similarly, the organization of 60 kb of a barley chromosome agreed with a gene-clustering model (11).

Contrary to previous observations, however, that genes in maize are usually separated from regions containing highly repetitive DNAs and rarely mixed within them (7–9), our results illustrate two patterns of gene distribution: individual genes amidst a sea of repetitive DNA, like 334B7.2, *adh1*, 334B7.4, and 334B7.9, separated by sizable blocks of retroelements, and clustered genes, like 334B7.5, 334B7.6, 334B7.7, and 334B7.8, occupying a space uninterrupted by highly repetitive DNAs. If the structural organization of the *adh1* region is a faithful representation of the general pattern of DNA organization in most of the maize genome, as suggested earlier (48), then different species might display different patterns. For large genomes, containing massive amounts of retrotransposon DNA, this study illustrates how important it is to know that large repetitive DNA blocks are not necessarily void of genes and that functional genes may be found interspersed within the repetitive DNAs.

The authors are indebted to Dr. Ben Bowen (Pioneer Hi-Bred) for providing EST Database access and some of the cDNA clones used for establishing relevant gene structures. This work was partially supported by Pioneer Hi-Bred CRA Program, by U.S. Department of Agriculture Grant 98-35300-6167 to Z.A., and by U.S. Department of Agriculture Grants 94-37300-0299 and 97-35300-4594 to J.L.B.

1. Hulbert, S. H., Richter, T. E., Axtell, J. D. & Bennetzen, J. L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4251–4255.
2. Ahn, S. & Tanksley, S. D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 7980–7984.
3. Moore, G., Foote, T., Helentjaris, T., Devos, K., Kurata, N. & Gale, M. (1995) *Trends Genet.* **11,** 81–82.
4. Devos, K. M. & Gale, M. D. (1997) *Plant Mol. Biol.* **35,** 3–15.
5. Bennetzen, J. L. & Freeling, M. (1997) *Genome Res.* **7,** 301–306.
6. Moore, G., Gale, M., Kurata, N. & Flavell, R. (1993) *Bio/Technology* **11,** 584–489.
7. Bernardi, G. & Bernardi, G. (1986) *J. Mol. Evol.* **24,** 1–11.
8. Barakat, A., Carels, N. & Bernardi, G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6857–6861.
9. Schmidt, T. & Heslop-Harrison, J. S. (1998) *Trends Plant Sci.* **3,** 195–199.
10. Wakimoto, B. T. (1998) *Cell* **93,** 321–324.
11. Panstruga, R., Busches, R., Piffanelli, P. & Schulze-Lefert, P. (1998) *Nucleic Acids Res.* **26,** 1056–1062.
12. Avramova, Z., Tikhonov, A., SanMiguel, P., Jin, Y. K., Liu, C., Woo, S. S., Wing, R. A. & Bennetzen, J. L. (1996) *Plant J.* **10,** 1163–1168.
13. Chen, M., SanMiguel, P., de Oliveira, A. C., Woo, S. S., Zhang, H., Wing, R. A. & Bennetzen, J. L. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 3431–3435.
14. Chen, M. S. & Bennetzen, J. L. (1996) *Plant Mol. Biol.* **32,** 999–1001.
15. Chen, M. S., SanMiguel, P. & Bennetzen, J. L. (1998) *Genetics* **148,** 435–443.
16. Laurie, D. A. & Bennett, M. D. (1985) *Heredity* **55,** 307–313.
17. Edwards, K. J., Thompson, H., Edwards, D., de Saizieu, A., Sparks, C., Thompson, J. A., Greenland, A. J., Eyers, M. & Schuch, W. (1992) *Plant Mol. Biol.* **19,** 299–308.
18. Woo, S. S., Jiang, J., Gill, B. S., Paterson, A. H. & Wing, R. A. (1994) *Nucleic Acids Res.* **22,** 4922–4931.
19. Del Sal, G., Manfioletti, G. & Scneider, C. (1988) *Nucleic Acids Res.* **16,** 9878.
20. Sambrook, J., Maniatis, T. & Fritsch, E. F. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
21. Bennetzen, J. L., Schrick, K., Springer, P. S., Brown, W. E. & SanMiguel, P. (1994) *Genome* **37,** 565–576.
22. Strathmann, M., Hamilton, B. A., Mayeda, C. A., Simon, M. I., Meyerowitz, E. M. & Palazzolo, M. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1247–1250.
23. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
24. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
25. Huang, X., Adams, M. D., Zhou, H. & Kerlavage, A. R. (1997) *Genomics* **46,** 37–45.
26. Springer, P. S., Edwards, K. J. & Bennetzen, J. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 863–867.
27. SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z. & Bennetzen, J. L. (1996) *Science* **274,** 765–768.
28. Avramova, Z., SanMiguel, P., Georgieva, E. & Bennetzen, J. L. (1995) *Plant Cell* **7,** 1667–1680.
29. SanMiguel, P., Gaut, B. S, Tikhonov, A., Nakajima, Y. & Bennetzen, J. L., (1998) *Nat. Genet.* **20,** 43–45.
30. White, S. E., Habera, L. F. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11792–11796.
31. Song, W.-Y., Pi, L.-Y., Bureau, T. E. & Ronald, P. C. (1998) *Mol. Gen. Genet.* **258,** 449–456.
32. Gaut, B. S. & Doebley, J. F. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6809–6814.
33. Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., *et al.* (1998) *Nature (London)* **391,** 485–488.
34. Tatusov, R. L., Mushegian, A. R., Bork, P., Brown, N. P., Hayes, W. S., Borodovsky, M., Rudd, K. E. & Koonin, E. V. (1996) *Curr. Biol.* **6,** 279–291.
35. Koonin, E. V., Mushegian, A. R., Galperin, M. Y. & Walker, D. R. (1997) *Mol. Microbiol.* **25,** 619–637.
36. Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B. & Brenner, S. (1996) *Trends Genet.* **12,** 145–150.
37. Koop, B. F. & Hood, L. (1994) *Nat. Genet.* **7,** 48–53.
38. Dunford, R. P., Kurata, N., Laurie, D. A., Money, T. A., Minobe, Y. & Moore, G. (1995) *Nucleic Acids Res.* **23,** 2724–2728.
39. Guimaraes, C. T., Sills, G. R. & Sobral, B. W. S. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14261–14266.
40. SanMiguel, P. & Bennetzen, J. L. (1998) *Ann. Bot. (London)* **82,** 37–44.
41. Berhan, A. M., Hulbert, S. H., Butler, L. G. & Bennetzen, J. L. (1993) *Theor. Appl. Genet.* **86,** 598–604.
42. Kilian, A., Kudrna, D. A., Kleinhofs, A., Yano, M., Kurata, N., Steffenson, B. & Sasaki, T. (1995) *Nucleic Acids Res.* **14,** 2729–2733.
43. Bennetzen, J. L., Schrick, K. M., Springer, P. S., Brown, W. E. & SanMiguel, P. (1994) *Genome* **37,** 565–576.
44. Cresse, A., Hulbert, S., Brown, W., Lucas, J. & Bennetzen, J. B (1995) *Genetics* **140,** 315–324.
45. Soltis, D. E. & Soltis, P. S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8089–8091.
46. Song, K., Lu, P., Tang, K. & Osborn, T. C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7719–7723.
47. Voytas, D. F. (1996) *Science* **274,** 257–261.
48. Edwards, K. J., Veuskens, J., Rawles, H., Daly, A. & Bennetzen, J. L (1996) *Genome* **39,** 811–817.

Alignment of SEQ ID NO: 1

```
AF123535
LOCUS       AF123535               160480 bp    DNA     linear   PLN 19-MAR-2000
DEFINITION  Zea mays alcohol dehydrogenase 1 (adh1) gene, adh1-F allele,
            complete cds.
ACCESSION   AF123535
VERSION     AF123535.1  GI:7262818
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 160480)
  AUTHORS   Tikhonov,A.P., SanMiguel,P.J., Nakajima,Y., Gorenstein,N.M.,
            Bennetzen,J.L. and Avramova,Z.
  TITLE     Colinearity and its exceptions in orthologous adh regions of maize
            and sorghum
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (13), 7409-7414 (1999)
   PUBMED   10377428
REFERENCE   2  (bases 1 to 160480)
  AUTHORS   SanMiguel,P.J., Tikhonov,A. and Bennetzen,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-1999) Biological Sciences, Purdue University,
            Hansen LSRB, Rm. 339, West Lafayette, IN 47907, USA
FEATURES             Location/Qualifiers
     source          1. .160480
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
                     /clone="YAC 334B7"
     mRNA            join(<17943. .17976,18512. .18648,18746. .18792,19208. .19533,
                     19620. .19702,19795. .19870,20216. .20277,20365. .20460,
                     20552. .20713,20814. .>20930)
                     /gene="adh1"
                     /product="alcohol dehydrogenase 1"
     gene            <17943. .>20930
                     /gene="adh1"
                     /allele="adh1-F"
     CDS             join(17943. .17976,18512. .18648,18746. .18792,19208. .19533,
                     19620. .19702,19795. .19870,20216. .20277,20365. .20460,
                     20552. .20713,20814. .20930)
                     /gene="adh1"
                     /note="ADH1-F"
                     /codon_start=1
                     /product="alcohol dehydrogenase 1"
                     /protein_id="AAF43977.1"
                     /db_xref="GI:7262819"
                     /translation="MATAGKVIKCKAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILFT
                     SLCHTDVYFWEAKGQTPVFPRIFGHEAGGIIESVGEGVTDVAPGDHVLPVFTGECKEC
                     AHCKSAESNMCDLLRINTDRGVMIADGKSRFSINGKPIYHFVGTSTFSEYTVMHVGCV
                     AKINPQAPLDKVCVLSCGISTGLGASINVAKPPKGSTVAVFGLGAVGLAAAAEGARIAG
                     ASRIIGVDLNPSRFEEARKFGCTEFVNPKDHNKPVQEVLAEMTNGGVDRSVECTGNIN
                     AMIQAFECVHDGWGVAVLVGVPHKDAEFKTHPMNFLNERTLKGTFFGNYKPRTDLPNV
                     VELYMKKELEVEKFITHSVPFAEINKAFDLMAKGEGIRCIIRMEN"
BASE COUNT    42638 a  37287 c  36880 g  43550 t    125 others
ORIGIN
```

June 22, 1999

```
  Query Match             18.4%;  Score 51.4;  DB 8;  Length 160480;
  Best Local Similarity   75.3%;  Pred. No. 5.1e-06;
  Matches   64;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy     196 tgagttgccttgcgcggaaggtgacgagtgaggtatctgtattccggaggcgtaagagtt 255
           || ||    || || ||  ||||| |||||||||| ||  ||  |||||||| || ||||
Db   57344 TGGGTCATTTTTCGTGGCGGGTGATGAGTGAGGTGTCCCTAACCCGGAGGCATAGGAGTT 57403

Qy     256 cctcggctcggtcgggcttgcccct 280
           ||||||||| ||||| || ||| ||
Db   57404 CCTCGGCTCAGTCGGCCTCGCCACT 57428
```

Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
OD and Molecular Weight calculations are for single-stranded DNA or RNA

Nucleotide base codes

GAG TTC CTC GGC TC

Reverse Complement Strand(5' to 3') is:

GAG CCG AGG AAC TC

Number of Fluorescent tags per strand:

0 6-FAM 0 TET 0 HEX 0 TAMRA DNA

Minimum base pairs required for single primer self-dimerization: 5

Minimum base pairs required for a hairpin : 4

Calculate | SWAP STRANDS | BLAST2 | Check Self-Complementarity

Physical Constants

Length: 14 bases

GC content: 64 %

Molecular Weight: 4371.8 4

1 ml of a sol'n with an Absorbance of 1 at 260 nm

is 7.571 microMolar 5 and contains 33.1 micrograms.

Melting Temperature (T_M) Calculations

1 43 °C (Basic)

2 50 °C (Salt Adjusted)

3 42 °C (Nearest Neighbor)

50 nM Primer

50 mM Salt (Na^+)

Thermodynamic Constants
Conditions: 1 M NaCl at 25°C at pH 7.

RlogK 33.404 cal/(°K*mol)

deltaH 116.5 Kcal/mol

deltaG 17.5 Kcal/mol

deltaS 302.8 cal/(°K*mol)

To use this calculator, you must be using Netscape 3.0 or later
or Internet Explorer version 3.0 or later, or another Javascript-capable browser
Self-Complementarity requires a 4.x browser. IE 5.0, Safari, and Mozilla supported.
This page was written in Javascript.
Extensively rewritten from 12/15/2000-12/19/2000 to isolate javascript Oligo object behaviors for teaching purposes.
This page may be freely distributed for any educational or non-commercial use.

About the Calculations

Thermodynamic Calculations

The nearest neighbor and thermodynamic calculations are done essentially as described by Breslauer *et al.*, *Proc. Nat. Acad. Sci.* **83**, 3746-50, 1986 (Abstract) but using the values published by Sugimoto *et al.*, *Nucl. Acids Res.* **24**, 4501-4505, 1996 (Abstract). This program assumes that the sequences are not symmetric and contain at least one G or C. The minimum length for the query sequence is 8.

The melting temperature calculations are based on the simple thermodynamic relationship between entropy, enthalpy, free energy and temperature, where

Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
OD and Molecular Weight calculations are for single-stranded DNA or RNA

Nucleotide base codes

CCG GAG GCG TAA GAG TTC CTC GGC TCG GTC GGG CTT GCC CCT

Reverse Complement Strand(5' to 3') is:

AGG GGC AAG CCC GAC CGA GCC GAG GAA CTC TTA CGC CTC CGG

Number of Fluorescent tags per strand:

0 6-FAM 0 TET 0 HEX 0 TAMRA DNA

Minimum base pairs required for single primer self-dimerization: 5

Minimum base pairs required for a hairpin : 4

Calculate | SWAP STRANDS | BLAST2 | Check Self-Complementarity

Physical Constants

Length: 42 bases

GC content: 69 %

Molecular Weight: 13056.3 [4]

1 ml of a sol'n with an Absorbance of 1 at 260 nm

is 2.408 microMolar [5] and contains 31.4 micrograms.

Melting Temperature (T_M) Calculations

[1] 77 °C (Basic)

[2] 76 °C (Salt Adjusted)

[3] 76 °C (Nearest Neighbor)

50 nM Primer

50 mM Salt (Na^+)

Thermodynamic Constants
Conditions: 1 M NaCl at 25°C at pH 7.

RlogK 33.404 cal/(°K*mol)

deltaH 386.9 Kcal/mol

deltaG 71.5 Kcal/mol

deltaS 1000.4 cal/(°K*mol)

To use this calculator, you must be using Netscape 3.0 or later
or Internet Explorer version 3.0 or later, or another Javascript-capable browser
Self-Complementarity requires a 4.x browser. IE 5.0, Safari, and Mozilla supported.
This page was written in Javascript.
Extensively rewritten from 12/15/2000-12/19/2000 to isolate javascript Oligo object behaviors for teaching purposes.
This page may be freely distributed for any educational or non-commercial use.

About the Calculations

Thermodynamic Calculations

The nearest neighbor and thermodynamic calculations are done essentially as described by Breslauer *et al.*, *Proc. Nat. Acad. Sci.* **83**, 3746-50, 1986 (Abstract) but using the values published by Sugimoto *et al.*, *Nucl. Acids Res.* **24**, 4501-4505, 1996 (Abstract). This program assumes that the sequences are not symmetric and contain at least one G or C. The minimum length for the query sequence is 8.

The melting temperature calculations are based on the simple thermodynamic relationship between entropy, enthalpy, free energy and temperature, where

$$\Delta H = \Delta G + T\Delta S$$

The change in entropy (order or a measure of the randomness of the oligonucleotide) and enthalpy (heat released or absorbed by the oligonucleotide) are directly calculated by summing the values for nucleotide pairs obtained by Breslauer *et al.*, *Proc. Nat. Acad. Sci.* **83**, 3746-50, 1986. The relationship between the free energy and the concentration of reactants and products at equilibrium is given by

$$\Delta G = RT \ln\left(\frac{[DNA \bullet primer]}{[DNA][primer]}\right)$$

Substituting the two equations gives us

$$\Delta H = T\Delta S + RT \ln\left(\frac{[DNA \bullet primer]}{[DNA][primer]}\right)$$

and solving for temperature T gives

$$T = \frac{\Delta H}{\Delta S + R \ln\left(\frac{[DNA \bullet primer]}{[DNA][primer]}\right)}$$

We can assume that the concentration of DNA and the concentration of the DNA-primer complex are equal, so this simplifies the equation considerably. It has been determined empirically that there is a 5 (3.4 by Sugimoto et al.) kcal free energy change during the transition from single stranded to B-form DNA. This is presumably a helix initiation energy. Finally, adding an adjustment for salt gives the equation that the Oligo Calculator uses:

$$T = \frac{\Delta H - 5\frac{kcal}{^\circ K\ mole}}{\Delta S + R \ln\left(\frac{1}{[primer]}\right)} + 16.6 \log_{10}([Na^+])$$

No adjustment constant for salt concentration is needed, since the various parameters were determined at 1 Molar NaCl, and the log of 1 is zero.

ASSUMPTIONS:
The thermodynamic calculations assume that the annealing occurs at pH 7.0. The melting temperature (Tm) calculations assume the sequences are not symmetric and contain at least one G or C. The oligonucleotide sequence should be at least 8 bases long to give reasonable Tms.

Basic Melting Temperature (Tm) Calculations

h b c e b g c ch

The two standard approximation calculations are used. For sequences less than 14 nucleotides the formula is

$$Tm = (wA+xT) * 2 + (yG+zC) * 4$$

where w,x,y,z are the number of the bases A,T,G,C in the sequence, respectively.

For sequences longer than 13 nucleotides, the equation used is

$$Tm = 64.9 + 41*(yG+zC-16.4)/(wA+xT+yG+zC)$$

ASSUMPTIONS:
Both equations assume that the annealing occurs under the standard conditions of 50 nM primer, 50 mM Na^+, and pH 7.0.

Salt Adjusted Melting Temperature (Tm) Calculations

A variation on two standard approximation calculations are used. For sequences less than 14 nucleotides the same formula as the basic calculation is use, with a salt concentration adjustment

$$Tm = (wA+xT)*2 + (yG+zC)*4 - 16.6*\log_{10}(0.050) + 16.6*\log_{10}([Na^+])$$

where w,x,y,z are the number of the bases A,T,G,C in the sequence, respectively.

The term $16.6*\log_{10}([Na^+])$ adjusts the Tm for changes in the salt concentration, and the term $\log_{10}(0.050)$ adjusts for the salt adjustment at 50 mM Na^+. Other monovalent and divalent salts will have an effect on the Tm of the oligonucleotide, but sodium ions are much more effective at forming salt bridges between DNA strands and therefore have the greatest effect in stabilizing double-stranded DNA.

For sequences longer than 13 nucleotides, the equation used is

$$Tm = 81.5 + (41 * (yG+zC)/(wA+xT+yG+zC)) - (500/(wA+xT+yG+zC)) + 16.6*\log_{10}([Na^+]) - 0.62F$$

This equation is most accurate for sequences longer than 50 nucleotides. It is valid for oligos longer than 50 nucleotides from pH 5 to 9. Symbols and salt adjustment term as above, with the term $(41 * (yG + zC-16.4)/(wA + xT + yG + zC))$ adjusting for G/C content and the term $(500/(wA + xT + yG + zC))$ adjusting for the length of the sequence, and F is the percent concentration of formamide.

For more information please see the reference:
Howley, P.M; Israel, M.F.; Law, M-F.; and M.A. Martin "A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus genomes." *J. Biol. Chem.* **254**, 4876-4883, 1979.

RNA melting temperatures

$$Tm = 79.8 + 18.5*\log_{10}([Na^+]) + (58.4 * (yG+zC)/(wA+xT+yG+zC)) + (11.8 * ((yG+zC)/(wA+xT+yG+zC))^2) - (820/(wA+xT+yG+zC))$$

Where yG+zC are the mole fractions of G and C in the oligo, L is the length of the shortest strand in the duplex.

ASSUMPTIONS:
These equations assume that the annealing occurs under the standard conditions of 50 nM primer and pH 7.0.

Molecular Weight Calculations

DNA Molecular Weight (for instance Oligonucleotides)
Molecular Weight = $(A_n \times 313.21) + (T_n \times 304.2) + (C_n \times 289.18) + (G_n \times 329.21) + 79.0$
A_n, T_n, C_n, and G_n are the number of each respective nucleotide within the polynucleotide. The addition of 79.0 gm/mole to the molecular weight takes into account the 5' monophosphate left by most restriction enzymes. No phosphate is present at the 5' end of strands made by primer extension.

RNA Molecular Weight (for instance from an RNA transcript)
Molecular Weight = (A_n x 329.21) + (U_n x 306.17) + (C_n x 305.18) + (G_n x 345.21) + 159.0
A_n, U_n, C_n, and G_n are the number of each respective nucleotide within the polynucleotide. Addition of 159.0 gm/mole to the molecular weight takes into account the 5' triphosphate.

OD Calculations

Molar Absorptivity values in 1/(Moles cm)

Residue	Moles^{-1} cm^{-1}	A_{max}(nm)	Molecular Weight (after protecting groups are removed)
Adenine (dAMP, Na salt)	15200	259	313.21
Guanine (dGMP, Na salt)	12010	253	329.21
Cytosine (dCMP, Na salt)	7050	271	289.18
Thymidine (dTMP, Na salt)	8400	267	304.2
RNA nucleotides			
Adenine (AMP, Na salt)	15400	259	329.21
Guanine (GMP, Na salt)	13700	253	345.21
Cytosine (CMP, Na salt)	9000	271	305.18
Uradine (UMP, Na salt)	10000	262	306.2
Other nucleotides			
6' FAM	20960		537.46
TET	16255		675.24
HEX	31580		744.13
TAMRA	31980		

Assume 1 OD of a standard 1ml solution, measured in a cuvette with a 1 cm pathlength.

6-FAM:

Chemical name:	**6-carboxyfluorescein**
Absorption wavelength maximum:	**495 nm**
Emission wavelength maximum:	**521 nm**
Molar Absorptivity at 260nm:	**20960** Moles^{-1} cm^{-1}

TET:

Chemical name:	**4, 7, 2', 7'-Tetrachloro-6-carboxyfluorescein**
Absorption wavelength maximum:	**519 nm**
Emission wavelength maximum:	**539 nm**
Molar Absorptivity at 260nm:	**16255** Moles^{-1} cm^{-1}

h b c e b g c ch

HEX:

Chemical name:	**4, 7, 2', 4', 5', 7'-Hexachloro-6-carboxyfluorescein**
Absorption wavelength maximum:	**537 nm**
Emission wavelength maximum:	**556 nm**
Molar Absorptivity at 260nm:	**31580** Moles^{-1} cm^{-1}

TAMRA:

Chemical name:	**N, N, N', N'-tetramethyl-6-carboxyrhodamine**
Absorption wavelength maximum:	**555 nm**
Emission wavelength maximum:	**580 nm**
Molar Absorptivity at 260nm:	**31980** Moles^{-1} cm^{-1}

Nucleotide base codes (IUPAC)

Symbol	nucleotide(s)				
A	adenine	**M**	A or C	**K**	G or T
C	cytosine	**R**	A or G	**V**	A or C or G; not T
G	guanine	**W**	A or T	**H**	A or C or T; not G
T	thymine in DNA; uracil in RNA	**S**	C or G	**D**	A or G or T; not C
		Y	C or T	**B**	C or G or T; not A
N	A or C or G or T				

Most recent version is available at URL: http://www.basic.northwestern.edu/biotools/oligocalc.html

---

The current version is the result of efforts by the following people:

Qing Cao, M.S. e-mail
Research Computing
Northwestern University Medical School
Chicago, IL 60611

Warren A. Kibbe, Ph.D. e-mail and PH entry.
Research Computing
Northwestern University Medical School
Chicago, IL 60611

Original code by Eugen Buehler
Research Support Facilities
Department of Molecular Genetics and Biochemistry
University of Pittsburgh School of Medicine

Monomer structures and molecular weights provided by Bob Somers, Ph.D.
e-mail
Sr. Applications Chemist
Glen Research Corporation
22825 Davis Drive

h b c e b g c c h

Sterling, VA 20164
http://www.glenres.com/

Uppercase/lowercase strand complementation problem described by Alexey Merz alexey@dartmouth.edu

Oligo Calculator version 3.03 (last modified by WAKibbe 02/12/2004)

h b c e b g c c h